

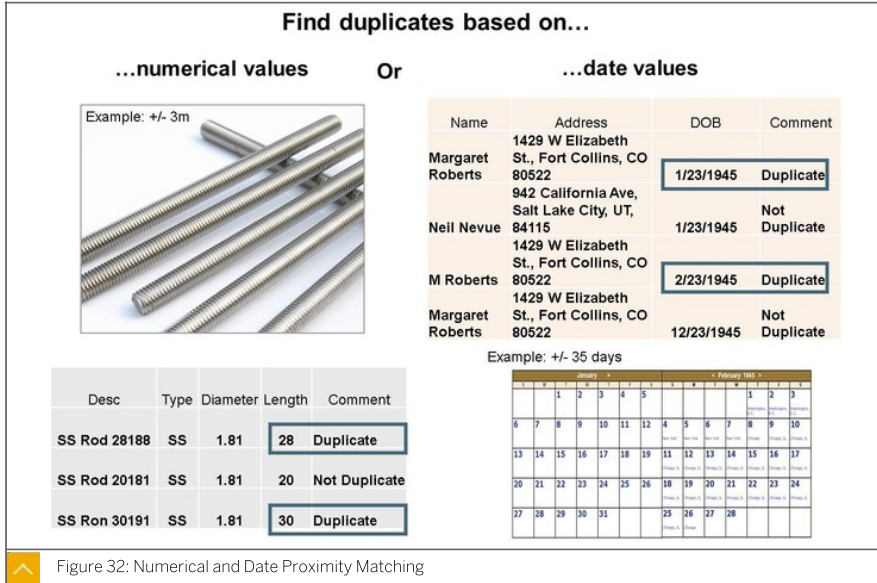


Figure 32: Numerical and Date Proximity Matching

Comparing Compound Family Name

This is an extension of the substring adjustment score option in the sense that it loosens some of the prerequisites. Not only can it find new matches, it can also boost the score for some existing matches. The *Approx substring* option works only on one field. If the family names are in separate fields, they need to be concatenated into one field. It may be useful to know how Adjustment scores are calculated may be useful, but not necessary for all to understand. Say there are two strings: Intl . and International. Without any adjustment scores, (Number of identical characters * 100)/Total number of characters = (8*100)/17 = 47. Now if the adjustment score is 80, it is applied to the characters that are not identical. Score = (Number of identical characters * 100) + (Number of nonidentical characters * 80)/Total number of characters = (8*100 + 9*80)/17 = 89.



Comparing Compound Family Names – Approximate Substring Matching

Approximate Substring Matching

- Mainly to support compound family names from Brazil, Mexico, and so on, but can be applied to non-name data. Data must be in *One input* field.
- Example CRUZ RODRIQUEZ and GARCIA CRUZ DE RDZ can match.
- Loosens some requirements of the *Substring matching* option:
 - First words do not have to match exactly, such as CRUZ and GARCIA.
 - The words that do match can utilize initials and abbreviations adjustments, such as RODRIGUEZ and RDZ.
 - Matching words have to be in the same order, but there can be other non-matching words around matching words, such as GARCIA, DE.